Daily pain prediction using smartphone speech recordings of patients with spine disease

Duey, A.H.; Rana, A.; Siddi, F.; Hussein, H.; Onnela, J.P.; Smith, T.R.

Daily Pain Prediction Using Smartphone Speech Recordings of Patients With Spine Disease

Akiro H. Duey, BS [‡§*], **Aakanksha Rana, PhD** [‡||*], **Francesca Siddi, MD** [‡¶*], **Helweh Hussein, MD** [‡], **Jukka-Pekka Onnela, DSc** [#], **Timothy R. Smith, MD, PhD, MPH** [‡**]

[‡]Department of Neurosurgery, Computational Neuroscience Outcomes Center, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [§]Icahn School of Medicine at Mount Sinai, New York, New York, USA; [||]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; [¶]Departments of Neurosurgery, Leiden University Medical Center, Leiden, The Netherlands; [#]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA; **Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

*Akiro H. Duey, Aakanksha Rana, and Francesca Siddi contributed equally to this work.

This work was presented as an oral poster on October 11, 2022 at the Congress of Neurological Surgeons Annual Meeting in San Francisco.

Correspondence: Timothy R. Smith, MD, PhD, MPH, Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, 60 Fenwood Rd, Boston, MA 02115, USA. Email: trsmith@bwh.harvard.edu.

Received, September 18, 2022; **Accepted,** February 02, 2023; **Published Online,** March 30, 2023.

BACKGROUND: Pain evaluation remains largely subjective in neurosurgical practice, but machine learning provides the potential for objective pain assessment tools.

OBJECTIVE: To predict daily pain levels using speech recordings from personal smartphones of a cohort of patients with diagnosed neurological spine disease.

METHODS: Patients with spine disease were enrolled through a general neurosurgical clinic with approval from the institutional ethics committee. At-home pain surveys and speech recordings were administered at regular intervals through the Beiwe smartphone application. Praat audio features were extracted from the speech recordings to be used as input to a K-nearest neighbors (KNN) machine learning model. The pain scores were transformed from a 0 to 10 scale to low and high pain for better discriminative capacity.

RESULTS: A total of 60 patients were enrolled, and 384 observations were used to train and test the prediction model. Using the KNN prediction model, an accuracy of 71% with a positive predictive value of 0.71 was achieved in classifying pain intensity into high and low. The model showed 0.71 precision for high pain and 0.70 precision for low pain. Recall of high pain was 0.74, and recall of low pain was 0.67. The overall F1 score was 0.73.

CONCLUSION: Our study uses a KNN to model the relationship between speech features and pain levels collected from personal smartphones of patients with spine disease. The proposed model is a stepping stone for the development of objective pain assessment in neurosurgery clinical practice.

KEY WORDS: Digital phenotyping, Speech analysis, Patient-reported outcome measures, Spine surgery, Machine learning, Smartphone

Neurosurgery 93:670–677, 2023

https://doi.org/10.1227/neu.0000000000002474

Pain due to spine disease can be debilitating and is a primary driver of spine surgery.[1] However, the subjective nature of pain makes its assessment and treatment challenging. The current gold standard for pain assessment in neurosurgical clinical practice includes methods such as Numeric Rating Scale (NRS) and visual analog scale; NRS is a numeric scale in which patients rate their pain on a scale between 0 and 10.[2,3] Although these pain measurement methods are proven simple pain assessment methods, they often require in-person clinic visits and rely on patient recollection.

ABBREVIATIONS: GPS, global positioning system; **KNN,** K-nearest neighbors; **MFCC,** Mel-frequency cepstral coefficient; **NRS,** Numeric Rating Scale; **PCA,** principal component analysis.

Supplemental digital content is available for this article at neurosurgery-online.com.

As a supplement to physical visits and recalls, remote health monitoring through modern digital devices such as smartphones and tablets has recently emerged to inform prevention and care management. Smartphone-based patient monitoring offers the possibility for clinicians to monitor their patients' health when they are at home interacting with the outside world. "Digital phenotyping," a term coined by Onnela and reported in Onnela et al,[4,5] is a technique which quantifies a patient's moment-by-moment phenotype driven by continuous data collection from mobile devices such as smartphones. Recent studies have found a correlation between pain and mobility as measured by smartphone global positioning system (GPS) data in patients with spine disease.[6,7] Digital phenotyping data, therefore, seem to hold great potential to track and assess pain which can supplement in-person clinic visits. Although pain is a sensation which includes elements of subjectivity, a digital phenotyping approach may better control for fluctuations in patient perception.

Prior research has shown that pain can modify speech in certain patients.[8] In addition, speech has been shown to provide valuable insight into aberrant brain activity for diagnosing depression,[9,10] schizophrenia,[11] and Alzheimer disease.[12] Therefore, speech data may represent a digital biomarker for pain which could supplement current pain assessment methods.

In neurosurgery research, the applicability of machine learning methods has increased substantially over the past few years.[13,14] Machine learning techniques in digital phenotyping data analysis are pertinent because of their ability to create models from large data sets for a variety of applications. Large-scale data streams collected in digital phenotyping have great potential; however, there are also significant challenges with implementing applications such as pain prediction in a clinical setting. The purpose of this study was to create a machine learning model using digital phenotyping speech data collected through a smartphone application to predict same-day pain levels in patients with spine disease.

METHODS

Patient Enrollment

This study was approved by the ethics committee at our institution (protocol number 2016P000095), and patient consent to participate in the study was obtained. The patients in this study were enrolled by the Neurosurgery Department at our institution between June 2017 and July 2019. The inclusion criteria for this study were patients with spine disease who presented to the neurosurgery department for treatment. Patients with a history of opioid abuse and patients undergoing multiple spine surgeries were excluded from the cohort. Not all patients underwent surgery before or during the study. Patient participation in the study included preoperative participation only, postoperative participation only, and both preoperative and postoperative participation. Inclusion criteria based on length of follow-up was not defined since each data point was treated independently. Participating patients were asked to download and install the Beiwe smartphone application (Onnela Lab LLC).[15] Patients who downloaded the Beiwe application but did not record a speech sample were excluded from the study.

The 2 sources of data for this study were pain surveys and speech recordings. The timing of speech and pain data collection was predetermined by the smartphone app without input from the authors. Pain surveys were prompted through smartphone notifications each day at 5 PM local time. The pain surveys were administered using the NRS pain scale. The survey text read, "Please rate your pain over the last 24 hours on a scale from 0 to 10; where 0 is no pain at all and 10 is the worst pain imaginable." Speech recordings were prompted once per week on Monday at 5 PM local time through smartphone notifications from the Beiwe application. Patients were asked to read aloud the first paragraph from "A Tale of Two Cities" by Charles Dickens. This passage was standardized across all patients and time points for each speech recording. Free speech responses, such as verbal expressions of how the patient was feeling, were not analyzed in this study. Data collection continued until the app was deleted, and all data were stored on a secure database. The data used in this study are not publicly available.

Data Preprocessing

Since patients did not always complete the pain surveys and speech recordings were only prompted once per week, the speech recordings and pain surveys were matched according to completion time. Only speech recordings with a pain survey completed within 24 hours before the recording were included in the analysis. If multiple pain surveys met the criteria, the pain survey completed closest to the time of the speech recording was selected. The speech recordings were manually screened to exclude all incomplete and accidental recordings. All unmatched pain surveys and speech recordings were excluded.

Feature Extraction

Speech recordings from the Beiwe application are provided in various formats. For compatibility with software packages, speech recordings

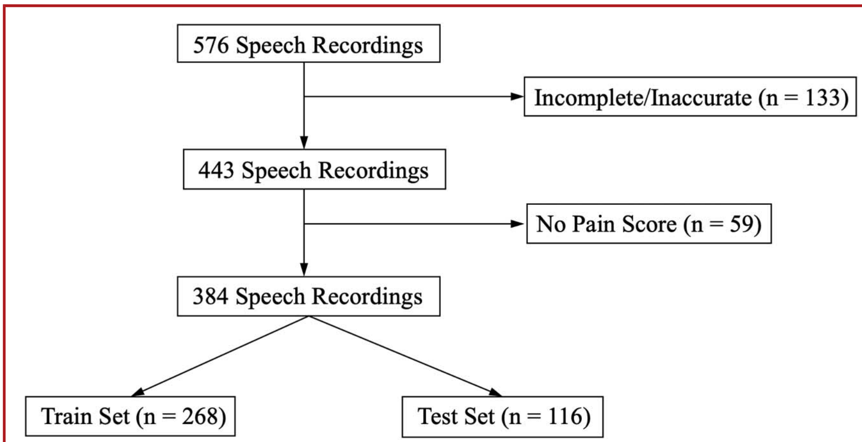

FIGURE 1. *Flow chart showing excluded patients and speech recordings.*

TABLE 1. Patient Characteristics

Characteristic	All patients (n = 60)
Age	
Mean (SD)	58.5 (12.7)
<50	14 (23.3%)
50-59	18 (30.0%)
60-69	17 (28.3%)
≥70	11 (18.3%)
Sex	
Male	26 (43.3%)
Female	34 (56.7%)
Race	
White	53 (88.3%)
Asian	2 (3.3%)
Hispanic	2 (3.3%)
Others	2 (3.3%)
N/A	3 (5.0%)
Spinal levels	
Cervical	13 (21.7%)
Thoracic	2 (3.3%)
Lumbar	37 (61.7%)
Sacral	2 (3.3%)
Mixed levels	6 (10.0%)
Primary diagnosis	
Central stenosis	22 (36.7%)
Herniated disk	12 (20.0%)
Foraminal stenosis	8 (13.3%)
Spondylolisthesis	8 (13.3%)
Scoliosis	4 (6.7%)
Others	6 (10.0%)

originally recorded in MP3 were converted to waveform audio file format format. Speech features were extracted from using the parselmouth Python library, an implementation of the Praat audio analysis software.[16] The list of extracted speech features consisted of meanF0Hz, stdevF0Hz, harmonic-to-noise ratio, localJitter, localabsoluteJitter, rapJitter, ppq5Jitter, localShimmer, localdbShimmer, apq3Shimmer, apq5Shimmer, apq11Shimmer, JitterPCA, and ShimmerPCA. F0 refers to the fundamental frequency, harmonic-to-noise ratio is the ratio between periodic and nonperiodic components, jitter describes frequency variation, and shimmer describes amplitude variation.[17] The Praat speech features are derived from the time domain. In addition, 13 Mel-frequency cepstral coefficient (MFCC) speech features from the frequency domain were extracted using the librosa Python library.[18]

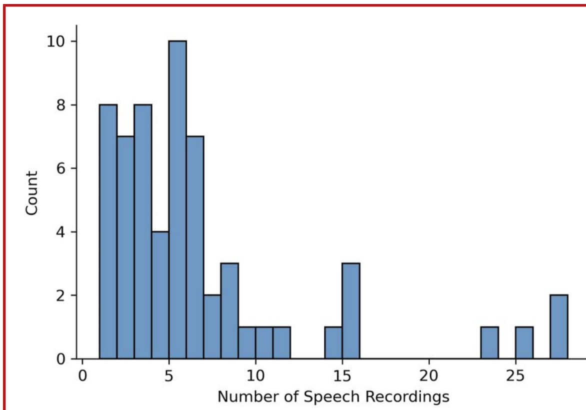

FIGURE 2. *Distribution of number of speech recordings per patient.*

Univariate Analysis

Linear mixed models were created to check for significant associations between speech features and pain scores. For each model, pain was the outcome variable, and each speech feature was treated as its own predictor. A random intercept was added for each patient to control for variations in baseline pain perception.

Model Training

To train the machine learning model, the data were prepared for 5-fold cross validation and eventually split into disjoint training (80%) and test sets (20%) such that patients in the test set were not included in the training data. To prevent model bias toward speech features with larger values, the train and test data were scaled using robust scaling before training.[19] Robust scaling is a statistical method which uses the interquartile range to scale data in a way that decreases the effect of outliers. Because of limitations of sample size, we created a binary classification model. Pain scores between 0 and 4 were classified as low pain, and pain scores between 5 and 10 were classified as high pain. These specific pain score groups were selected to provide equal samples of "high" and "low" pain scores.

K-Nearest Neighbors–Based Prediction Model

A K-nearest neighbor (KNN) classifier was used to classify speech recordings into low or high pain categories.[20] We selected this algorithm because of its simplicity and clear interpretation. KNN is one of many different machine learning algorithms commonly used in medical research. The algorithm functions by plotting the training data in the n-dimensional space and classifying each test data point according to the K-nearest training data points. Logistic regression and random forest models were also trained. The Scikit-learn python library was used to build and train the machine learning models.[19] All models were trained using the following computational hardware: Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz, Architecture x86_64, Operating system Ubuntu 20.04.1 LTS, CPU(s): 8.

TABLE 2. Speech Recordings and Pain Scores

Characteristic	All patients (n = 60)
Speech samples per patient	
Median (IQR)	5 (2.5-7.0)
Range	1-28
Pain scores	
Mean (SD)	4.6 (2.7)
Range	0-10

RESULTS

Patient Demographics

After applying exclusion criteria, 384 speech recordings from 60 different patients were used to train and test the KNN model (Figure 1). The mean age was 58.5 years, and 26 (43.3%) patients were male. The median follow-up interval was 33 days (range 0-187 days). Fifty-three (88.3%) patients were white. A majority of patients had lumbar spine disease (61.7%) while cervical spine disease (21.7%) was the second most common location of spine disease (Table 1). Patient spine diagnoses included central stenosis in 22 (36.7%) patients, herniated disk in 12 (20.0%), foraminal stenosis in 8 (13.3%), spondylolisthesis in 8 (13.3%), and scoliosis in 4 (6.7%). Fracture, epidural mass, sacroiliac joint pathology, and discitis were considered "other" diagnoses.

Thirty-six (60.0%) of the 60 patients in the study underwent surgery. Of the patients undergoing surgery, 27 patients (75.0%) had at least 1 speech recording before surgery, and 26 patients (72.2%) had at least 1 speech recording after surgery. Seventeen patients (47.2%) had at least 1 speech recording both before and after surgery. Among patients with at least 1 speech recording after surgery, the median follow-up from the date of surgery was 47 days (range 0-273 days).

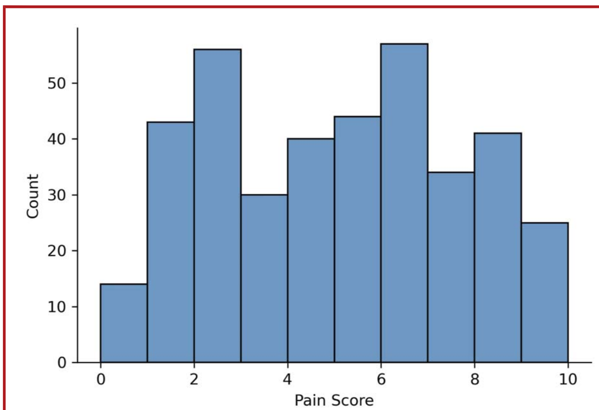

FIGURE 3. *Distribution of pain scores.*

Pain and Speech Data

The median number of audio files per patient was 5 (range 1-28, Figure 2), and the mean pain score was 4.6 ± 2.7 (Table 2). A histogram of pain scores is provided in Figure 3. The mean audio recording length was 43.3 ± 10.1 seconds. After converting the pain scores into low pain and high pain, there were 183 speech samples with low pain scores and 201 samples with high pain scores.

Univariate Analysis

The results from our linear mixed model analysis showed that none of the Praat speech features were significantly associated with pain scores (Table 3). MFCC 2 (Coef: 0.552; 95% CI: [0.254, 0.850]; $P < .001$) and MFCC 12 (Coef: −0.286; 95% CI: [−0.538, −0.034]; $P = .026$) were significantly associated with pain scores (**Supplemental Digital Content 1**, http://links.lww.com/NEU/D762).

Model Performance

There were 268 speech recordings in the training set and 116 speech recordings in the test set. In the test set, 55 (47.4%) samples were low pain and 61 (52.6%) were high pain. The KNN prediction accuracy on the test data set was 71%. A confusion matrix is presented in Figure 4, which shows a precision of 0.71 and recall of 0.74 for high pain. For low pain, the precision was 0.70 and recall was 0.67. The overall F1 score was 0.73 (Table 4). Logistic regression and random forest model results are provided in **Supplemental Digital Content 2 & 3**, http://links.lww.com/NEU/D763 and http://links.lww.com/NEU/D764. KNN model results using MFCC features as predictors are provided in **Supplemental Digital Content 4**, http://links.lww.com/NEU/D765.

DISCUSSION

Spine disease remains a common pathology in neurosurgical practice and often leads to significant and multifactorial pain. Although the current pain assessment methods have been validated in clinical practice, they are subject to patient recollection biases and require in-person clinic visits. Continuous and remote pain monitoring through smartphones and other mobile devices would be invaluable to neurosurgeons in their patient evaluations. Using digital phenotyping data, it may be possible to supplement the current pain assessment tools with timely pain assessment in a home setting which controls for variations in patient perceptions of their pain. The purpose of our study was to build a data-driven machine learning model to predict same-day pain levels using speech data collected from personal smartphones of patients with spine disease.

Pain recollection is multifaceted and depends on severity, chronicity, and timing of the pain. Pain intensity is often cited as the most critical element in pain perception,[21] and severe pain can cause decreased mobility, sleep deprivation, medication dependency, and anxiety.[22,23] Furthermore, chronic pain can be complex and often changes with time. Therefore, it can be difficult for patients to communicate their pain experiences to neurosurgeons

TABLE 3. Linear Mixed Model Results Showing Associations Between Praat Speech Features and Pain Scores With a Random Intercept for Each Patient

Audio feature	Coefficient	95% CI	*P* value
meanF0Hz	0.038	(−0.188, 0.265)	.739
stdevF0Hz	0.005	(−0.197, 0.206)	.964
HNR	0.159	(−0.071, 0.388)	.176
localJitter	−0.086	(−0.302, 0.130)	.437
localabsoluteJitter	−0.122	(−0.412, 0.169)	.412
rapJitter	−0.096	(−0.310, 0.118)	.380
ppq5Jitter	−0.053	(−0.272, 0.165)	.632
ddpJitter	−0.096	(−0.310, 0.118)	.380
localShimmer	−0.112	(−0.327, 0.103)	.308
localdbShimmer	−0.084	(−0.299, 0.131)	.443
apq3Shimmer	−0.108	(−0.326, 0.111)	.334
apq5Shimmer	−0.067	(−0.277, 0.144)	.533
apq11Shimmer	−0.055	(−0.257, 0.147)	.593
ddaShimmer	−0.108	(−0.326, 0.111)	.334
JitterPCA	−0.101	(−0.323, 0.120)	.369
ShimmerPCA	−0.003	(−0.206, 0.199)	.975

HNR, harmonic-to-noise ratio; PCA, principal component analysis.

TABLE 4. Results of K-Nearest Neighbor Pain Prediction Model Using the Test Data Set

Testing characteristics	Model
Low pain scores	55
High pain scores	61
True positives	45
True negatives	37
False positives	18
False negatives	16
Sensitivity (true-positive rate)	0.74
Specificity (true-negative rate)	0.67
Positive predictive value	0.71
Negative predictive value	0.67
Accuracy	0.71
F1 score	0.73

between follow-up visits. In addition, self-reported pain measurement can be biased by pain intensity and time of day.[24] On the other hand, digital phenotyping provides insight into a patient's daily functioning by (1) taking advantage of the ubiquity and familiarity of smartphones and (2) ease associated with data procurement.[4,25]

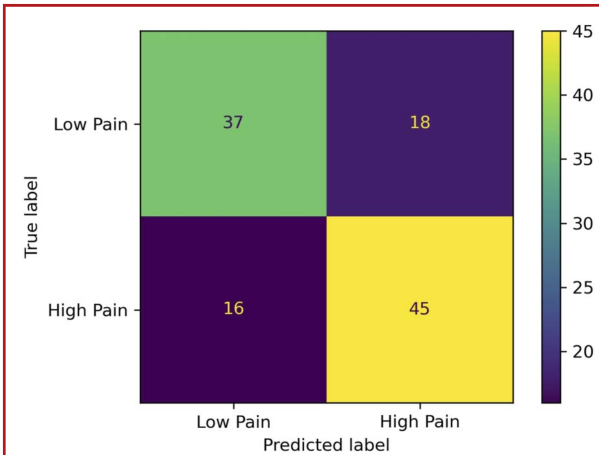

FIGURE 4. *Confusion matrix of K-nearest neighbor pain prediction model.*

Continuous patient monitoring using patient smartphones within neurosurgery has been studied previously. Digital phenotyping studies by Cote et al and Boaro et al used GPS data from the Beiwe smartphone application to study pain and patient reported outcomes in patients with spine disease. In 2019, Cote et al[6] implemented a linear mixed model approach which found that higher pain correlated with less mobility, as measured by GPS summary statistics. More recently in 2021, Boaro et al[7] showed that GPS summary statistics were significantly correlated with visual analog scale, Oswestry Disability Index, and Patient-Reported Outcome Measurement Information System 10 mental and physical scores. These studies together prove that digital phenotyping data collected from personal smartphones can reveal patient well-being at home. Our study adds to this literature by providing a framework to objectively and continuously assess pain due to spine disease in the home setting.

The human body likely exhibits pain signals through various media including electrical signals, facial expression, and speech. Previous studies have used speech features to detect depression,[9,10] Alzheimer disease,[12] and schizophrenia.[11] In an observational study by Di Matteo et al,[26] ambient speech was recorded from smartphones of patients with depression and anxiety and correlated with clinical symptoms. In addition, Laguarta et al[27] presented a model which was able to discriminate between patients with and without COVID-19 through analysis of a large database of cough audio recordings from more than 5000 participants. These studies provide evidence that speech is a physiologic conduit for diagnosing many human diseases. There is potential to create speech

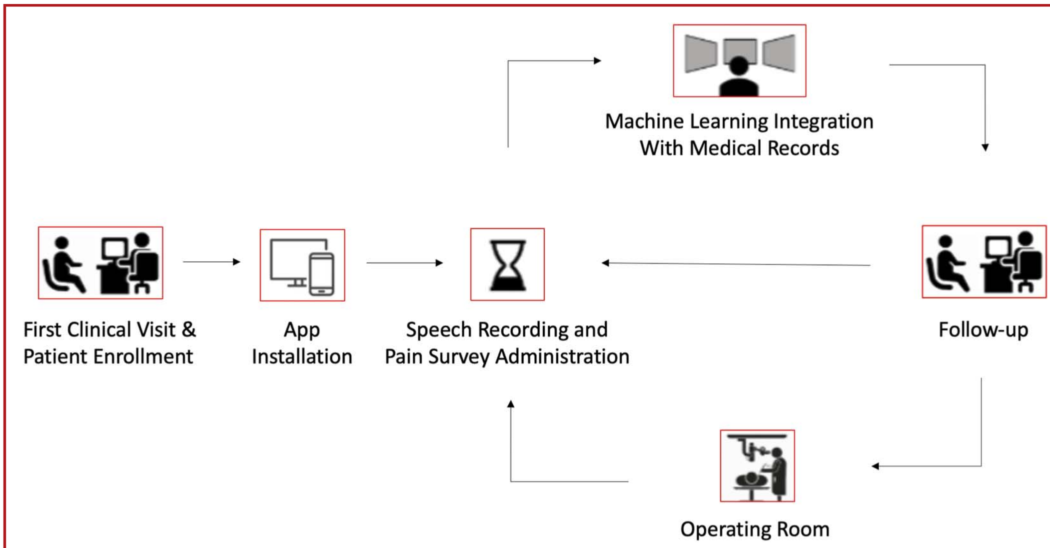

FIGURE 5. *Suggested digital phenotyping workflow with machine learning integration into medical records. During the first clinical visit, patient consent is acquired and the digital phenotyping application is installed onto the patient's smartphone. From this moment, digital phenotyping data are collected over time in the patient's home environment (surveys, voice recordings, mobility data, and sociability data). The digital phenotyping data and machine learning model enriche electronic medical records with its estimation and, health care professionals are able to follow the patient over time to provide a well-informed treatment plan.*

biomarkers which represent meaningful phenotypes of pain based on data collected from mobile devices. However, it is unclear which speech features would be best suited for understanding pain.

In our study, we used audio features from the Praat software because of their simplicity and wide adoption in speech research.[28-30] We also tested MFCC features with our machine learning model. Although MFCC 2 and 12 were significantly associated with pain scores in the univariate mixed model analysis, this did not translate to an improved KNN model accuracy compared with the KNN model trained with Praat features. MFCCs are commonly used audio features in speech analysis which represent a transformation of audio signals that mimics how the human ear perceives sound.[31,32] Researchers have developed other audio features to study human speech. OpenSmile is a common speech analysis software which includes some aspects of the MFCC features in addition to many other features.[33,34] This diversity of speech features should be further explored to discover those which are influential in understanding pain.

Several machine learning techniques have gained recent attention to automatically analyze biomarkers and predict real-time pain, including KNN, support vector machines, and tree-based approaches such as random forest algorithms. Because of their ability to learn patterns from massive data sets, machine learning techniques are powerful tools to harness clinically relevant signatures and build a predictive model. In 2021, Kong et al[35] published a study in which the authors built a random forest machine learning model to predict real-time pain using electrodermal activity from a calibrated wrist device. Similarly, Hasan et al[36] implemented a support vector machine classification model with facial recognition technology to predict pain. However, little has been done towards training machine learning models that use audio biomarkers to predict pain.

Motivated by these previous studies, our study explores machine learning techniques to harness the speech features as predictors. Given our relatively small data set, we used a KNN machine learning model which we determined to be the best fit for our speech data. These days deep learning frameworks include convolutional neural networks, long-short term memory networks, and transformers which can learn the complex time series relationship of speech signals.[37,38] Such models may better handle heterogeneous populations and improve discriminative capacity at multiple pain levels. Although we experimentally explored other models including random forests, support vector machines, and artificial neural networks, these more powerful algorithms typically perform better with larger data sets.

One advantage of a digital phenotyping approach to pain assessment is a potential reduction in administrative fees associated with survey administration. To implement our technology in a clinic setting, we believe that integration with the electronic medical record would be essential. Furthermore, it is important to standardize data collection across institutions and simplify participation for patients. Our suggested workflow includes an example of how our tool might be implemented in clinical practice. Between visits, patients would be asked to fill out regular pain surveys and provide sample voice recordings using their smartphone. These pain surveys and voice recordings would be combined with other digital phenotyping data sources in a comprehensive machine learning model to provide an estimation of pain while correcting for variations in patient perceptions over time. Next, the model would directly upload pain estimates to the electronic medical record and the patient's survey responses for direct comparison. These data would be used by both the physician and patient at their next visit while discussing an appropriate treatment plan. Our proposed digital phenotyping workflow has been illustrated in Figure 5.

Limitations

Although we believe that this study is a valuable addition to the literature, there are a few limitations that we would like to acknowledge. Our study was limited to spine patients at a single institution and may not generalize to other neurosurgery clinics, and our limited sample size of 60 patients total is not sufficient to build a generalizable model. A multicenter study with patients from different disciplines could reach more generalizable results. Our pain prediction model was based on self-reported pain scores, and, therefore, predictions are based on patient perception of pain rather than on an objective measure. As the time interval between pain surveys and speech recordings ranged up to 24 hours, speech features may not capture brief pain episodes occurring at the time of the survey and thus our study is better suited for the assessment of chronic pain. Moreover, there was significant variation in the length of follow-up and participation frequency of each patient. Consequently, our data were not independent with some patients contributing multiple speech samples. Although we chose to use a KNN model for simplicity and interpretability, a model which accounts for correlated data may be a better choice in the long term. With a larger data set, we can optimize predictive accuracy by analyzing more complex speech features with deep learning approaches.

CONCLUSION

This study offers the opportunity for practical objective pain assessment of neurosurgical spine patients in a home setting. There is limited research investigating associations between smartphone speech data and pain in neurosurgical patients. Herein, we propose a machine learning–based methodology to quantify pain using speech data from smartphones and self-reported pain surveys in a cohort of patients with spine disease. Using our prediction model as a baseline, future models can improve on our framework to better evaluate pain levels in patients with spine disease.

Funding

This study did not receive any funding or financial support.

Disclosures

Timothy Smith and Jukka-Pekka Onnela are cofounders of Phebe Health, a newly established company that operates in digital phenotyping. The other authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Lurie J, Tomkins-Lane C. Management of lumbar spinal stenosis. *BMJ.* 2016;352: h6234.
2. Younger J, McCue R, Mackey S. Pain outcomes: a brief review of instruments and techniques. *Curr Pain Headache Rep.* 2009;13(1):39-43.
3. Bendinger T, Plunkett N. Measurement in pain medicine. *BJA Educ.* 2016;16(9): 310-315.
4. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology.* 2016;41(7): 1691-1696.
5. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Ment Health.* 2016;3(2):e16.
6. Cote DJ, Barnett I, Onnela JP, Smith TR. Digital phenotyping in patients with spine disease: a novel approach to quantifying mobility and quality of life. *World Neurosurg.* 2019;126:e241-249.
7. Boaro A, Leung J, Reeder HT, et al. Smartphone GPS signatures of patients undergoing spine surgery correlate with mobility and current gold standard outcome measures. *J Neurosurg Spine.* 2021;35(6):796-806.
8. Rowbotham S, Wardy AJ, Lloyd DM, Wearden A, Holler J. Increased pain intensity is associated with greater verbal communication difficulty and increased production of speech and co-speech gestures. *PLoS One.* 2014;9(10):e110779.
9. Taguchi T, Tachikawa H, Nemoto K, et al. Major depressive disorder discrimination using vocal acoustic features. *J Affect Disord.* 2018;225:214-220.
10. Wang J, Zhang L, Liu T, Pan W, Hu B, Zhu T. Acoustic differences between healthy and depressed people: a cross-situation study. *BMC Psychiatry.* 2019;19(1):300.
11. Mota NB, Copelli M, Ribeiro S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *NPJ Schizophrenia.* 2017;3(1):18.
12. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis.* 2015;49(2):407-422.
13. Dagi TF, Barker FG, Glass J. Machine learning and artificial intelligence in neurosurgery: status, prospects, and challenges. *Neurosurgery.* 2021;89(2):133-142.
14. Senders JT, Zaki MM, Karhade AV, et al. An introduction and overview of machine learning in neurosurgical care. *Acta Neurochir.* 2018;160(1):29-38.
15. Onnela JP, Dixon C, Griffin K, et al. Beiwe: a data collection platform for high-throughput digital phenotyping. *J Open Source Softw.* 2021;6(68):3417.
16. Jadoul Y, Thompson B, de Boer B. Introducing parselmouth: a Python interface to Praat. *J Phon.* 2018;71:1-15.
17. Teixeira JP, Oliveira C, Lopes C. Vocal acoustic analysis—jitter, shimmer and HNR parameters. *Proced Technol.* 2013;9:1112-1122.
18. McFee B, Metsai A, McVicar M, et al. *Librosa/librosa: 0.9.2.*
19. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12(85):2825-2830.
20. Mucherino A, Papajorgji PJ, Pardalos PM. Nearest neighbor classification. *Data Mining Agri.* 2009:83-106.
21. Caraceni A. Evaluation and assessment of cancer pain and cancer pain treatment. *Acta Anaesthesiol Scand.* 2001;45(9):1067-1075.
22. Brennan F, Carr DB, Cousins M. Pain management: a fundamental human right. *Anesth Analg.* 2007;105(1):205-221.
23. Sinatra R. Causes and consequences of inadequate management of acute pain. *Pain Med.* 2010;11(12):1859-1871.
24. Jensen MP, Mardekian J, Lakshminarayanan M, Boye ME. Validity of 24-h recall ratings of pain severity: biasing effects of "Peak" and "End" pain. *Pain.* 2008; 137(2):422-427.
25. Torous J, Staples P, Onnela JP. Realizing the potential of mobile mental health: new methods for new data in psychiatry. *Curr Psychiatry Rep.* 2015;17(8):61.
26. Di Matteo D, Wang W, Fotinos K, et al. Smartphone-detected ambient speech and self-reported measures of anxiety and depression: exploratory observational study. *JMIR Form Res.* 2021;5(1):e22723.
27. Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open J Eng Med Biol.* 2020;1:275-281.
28. Gorris C, Ricci MA, Vanoni F, et al. Acoustic analysis of normal voice patterns in Italian adults by using Praat. *J Voice.* 2020;34(6):961.e9-961.e18.
29. Yilmaz A, Sarac ET, Aydinli FE, Yildizgoren MT, Okuyucu EE, Serarslan Y. Investigating the effect of STN-DBS stimulation and different frequency settings on the acoustic-articulatory features of vowels. *Neurol Sci.* 2018;39(10):1683-1689.

30. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. *Laryngoscope Invest Otolaryngol.* 2020;5(1):96-116.
31. Dash TK, Mishra S, Panda G, Satapathy SC. Detection of COVID-19 from speech signal using bio-inspired based cepstral features. *Pattern Recognit.* 2021;117:107999.
32. Eskidere Ö, Gürhanlı A. Voice disorder classification based on multitaper Mel frequency cepstral coefficients features. *Comput Math Methods Med.* 2015;2015:1-12.
33. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. ACM Conferences. Accessed July 27, 2022. https://dl.acm.org/doi/10.1145/1873951.1874246
34. de Boer JN, Voppel AE, Brederoo SG, et al. Acoustic speech markers for schizophrenia-spectrum disorders: a diagnostic and symptom-recognition tool. *Psychol Med.* 2023;53:1302-1312.
35. Kong Y, Posada-Quintero HF, Chon KH. Real-time high-level acute pain detection using a smartphone and a wrist-worn electrodermal activity sensor. *Sensors.* 2021;21(12):3956.
36. Hasan MK, Ahsan GMT, Ahamed SI, Love R, Salim R. Pain level detection from facial image captured by smartphone. *J Inf Process.* 2016;24(4):598-608.
37. Liang YC, Wijaya I, Yang MT, Cuevas Juarez JR, Chang HT. Deep learning for infant cry recognition. *Int J Environ Res Public Health.* 2022;19(10):6311.
38. Wang F. Machine learning for predicting rare clinical outcomes-finding needles in a Haystack. *JAMA Netw Open.* 2021;4(5):e2110738.

Supplemental digital content *is available for this article at neurosurgery-online.com.*

Supplemental Digital Content 1. Table. Linear mixed model results using Mel-frequency cepstral coefficient (MFCC) audio features with pain scores as the model output. MFCC 2 and 12 were significantly associated with pain scores.

Supplemental Digital Content 2. Table. Logistic regression model results with pain level (low, high) as the outcome variable and Praat audio features as the predictor variables.

Supplemental Digital Content 3. Table. Random forest classifier model results with pain level (low and high) as the outcome variable and Praat audio features as the predictor variables.

Supplemental Digital Content 4. Table. KNN model results with pain level (low and high) as the outcome variable and Mel-frequency cepstral coefficient (MFCC) audio features as the predictor variables.